AD-781 295

JACOB MARSCHAK AND OTHERS--
PERSONAL PROBABILITIES OF PROBABILITIES

Jacob Marschak

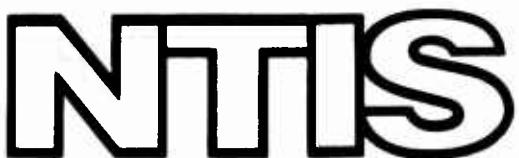California University

AD-781295

DOCUMENT CONTROL DATA · R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY *(Corporate author)*	2a REPORT SECURITY CLASSIFICATION
Western Management Science Institute University of California, Los Angeles	Unclassified
	2b GROUP

3 REPORT TITLE

JACOB MARSCHAK AND OTHERS--PERSONAL PROBABILITIES OF PROBABILITIES

4 DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Working Paper

5 AUTHOR(S) *(Last name, first name, initial)*

Marschak, Jacob

6 REPORT DATE	7a TOTAL NO. OF PAGES	7b NO OF REFS
May, 1974	47	29

8a CONTRACT OR GRANT NO	9a ORIGINATOR'S REPORT NUMBER(S)
N00014-69-A-0200-4005	
b PROJECT NO.	Working Paper No. 215
c	9b. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
d	

10 AVAILABILITY/LIMITATION NOTICES

11 SUPPLEMENTARY NOTES	12 SPONSORING MILITARY ACTIVITY

13 ABSTRACT

By definition, the subjective probability distribution of a random event is revealed by the ("rational") subject's choice between bets--a view expressed by F. Ramsey, B. De Finetti, L.J. Savage and traceable to E. Borel and, it can be argued, to T. Bayes. Since hypotheses are not observable events, no bet can be made, and paid off, on a hypothesis. The subjective probability distribution of hypotheses (or of a parameter, as in the current "Bayesian" statistical literature) is therefore a figure of speech, an "as if," justifiable in the limit. Given a long sequence of previous observations, the subjective posterior probabilities of events still to be observed are derived by using a mathematical expression that would approximate the subjective probability distribution of hypotheses, <u>if</u> these could be bet on. This position was taken by most, but not all, respondents to a "Round Robin" initiated by J. Marschak after M.H. DeGroot's talk on Stopping Rules presented at the UCLA Interdisciplinary Colloquium on Mathematics in Behavioral Sciences. Other participants: K. Borch, H. Chernoff, R. Dorfman, W. Edwards, G. Graves, K. Miyasawa, P. Randolph, L.J. Savage, R. Schlaifer, R.L. Winkler.
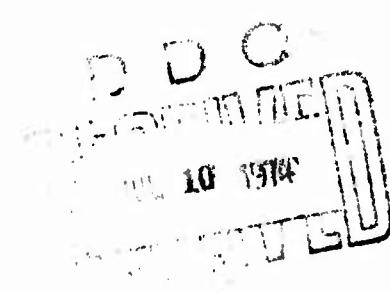
DD FORM 1 JAN 64 **1473**   0101-807-6800

ii

WESTERN MANAGEMENT SCIENCE INSTITUTE

University of California, Los Angeles

Working Paper No. 215

JACOB MARSCHAK AND OTHERS

PERSONAL PROBABILITIES OF PROBABILITIES

Jacob Marschak

May, 1974

JACOB MARSCHAK AND OTHERS

PERSONAL PROBABILITIES OF PROBABILITIES

Introduction

The discussion that follows originated at one of the bi-weekly
sessions of the Interdisciplinary Colloquium on Mathematics in
Behavioral Science, University of California at Los Angeles
(J. Marschak, chairman).[1] A paper by Morris H. DeGroot, on
Stopping Rules, elicited comments on the following question.
Suppose all probabilities are defined as "subjective", "personal"--
i.e., as being revealed by a "rational", "consistent" decision-
maker's choices under uncertainty: if he chooses to bet on one
rather than on another event, the former event is said to be the
subjectively more probable one. Moreover, a few rather plausible
quasi-logical postulates of "rationality" of choices imply that
such "subjective probabilities" have indeed the properties of a
mathematical "probability measure."-- Our question is: what
meaning, if any, can be assigned to the probability of a hypo-
thesis, law, theory that is itself a probability distribution so
that its falsity or truth is not, in general, an observable event
on which bets can be made and paid off. If an urn is sealed,
bets can be taken, both before and after some drawings were made,
on what the outcome of subsequent drawings will be. For these
outcomes will be observed and the bets paid off. But no bets
can be paid off on the content of the urn itself unless it is
unsealed. Most laws, theories, hypotheses are urns sealed
forever. Statisticians who speak of the prior and posterior

distribution of a statistical parameter use, in fact, a figure of speech which deserves clarification. J. Marschak's comments were therefore circulated to a number of workers in this field. Several have answered. A "Round Robin" (the late L.J. Savage's favorite term) resulted. Its publication was permitted by all contributors. They are (alphabetically)

Karl Borch	Koichi Miyasawa
Herman Chernoff	Paul Randolph
Morris H. DeGroot	L.J. Savage
Robert Dorfman	Robert Schlaifer
Ward Edwards	Robert L. Winkler
T.S. Ferguson	

Students of the fundamentals of decision theory and statistical inference know, and it is partly shown in the attached Bibliography, that most of the contributors have continued to work in the field.

As the reader will see, most respondents agreed that, to assign probability to the truth of a probabilistic theory is to use a figure of speech, a "as if", that is justified <u>in the limit</u>. For, given a sufficiently long sequence of observed events, one can compute approximate subjective probabilities of events still to be observed, by using as an intermediate step a mathematical expression which could be called the subjective probability distribution of hypotheses if it were possible to bet on a hypothesis. A few of the respondents, however, denied the existence of the problem itself, by permitting to associate the term "subjective probability" with a subject's naming a number, and not necessarily with his choosing between decisions.

The topic of the present discussion is far from outdated.
Bruno DeFinetti has treated rather recently (1971) the
"Probability of a Theory and Probabilities of Facts", with ex-
tensive reference to an instructive example provided by I .J.
Good (1969).

I may be permitted to add two remarks of my own. One
concerns a question of history and of definition. It was raised
by L.J. Savage in response to Item 7. of my comments. Did
Thomas Bayes (1763) interpret probabilities as subjective ones,
revealed by the bets of a rational decision-maker? Was The
Reverend a "personalist", anticipating Borel (1924), Ramsey
(1924-28), DeFinetti (1937), Savage (1954)? To-day's term
"Bayesian statistics" seems mostly to denote the use of prior and
the derivation of posterior probabilities (by "Bayes' Theorem");
in addition, viewing himself as a decision-maker, the "Bayesian
statistician" is supposed to concern himself with the posterior
expectation of the "loss" (the negative of the economists' "gross
payoff" or "benefit") and cost. But, to be truly "Bayesian", does
he not have, in addition, to interpret his prior and posterior
probabilities as subjective ones? It is not very important, of
course, what labels we use, provided we agree on their meaning.
Yet there may be some advantage if, in addition, we agree with
history when the label is a historical name.[2] Let me quote
"Definition 5" of Bayes:

> "The probability of any event is the ratio between the
> value at which an expectation depending on the happening
> of the event ought to be computed (J.M.'s italics), and
> the value of the thing expected upon its happening".

Does not the "ought to" indicate a <u>norm of behavior</u>, stating
what we would now call "consistent", "rational" behavior? To
the value (u dollars, say) that you would gain if the uncertain
event happens, there "ought to" correspond in your mind a smaller
but sure value (c dollars) such that you are indifferent between
gaining u upon the happening of the uncertain event and gaining
c with certainty. Hence there ought to be in your mind also a
ratio of these two numbers, c/u = p (say), which Bayes calls
the probability of that uncertain event. This argues, I think
for my ascribing him the personalistic view. Note also that he
calls our c the "value at which an expectation depending on the
event ought to be computed". This actually agrees with the
modern use of ("mathematical") expectation: for Bayes ob-
viously assumes that the payoff if that uncertain event does not
happen is = 0, so that indeed, if p = c/u then up + 0·(1-p) = c.
This interpretation of Bayes' term "expectation" is confirmed for
the case when the bettor's loss > 0, in Bayes' "Proposition" 2:

> "If a person has an expectation depending on the happening
> of an event, the probability of the event is to the pro-
> bability of its failure as his loss is if it fails to his
> gain if it happens".

Here "the person (who) has an expectation" considers a fair bet.
That is, denoting his gain and loss (both uncertain) by <u>u</u> and ℓ,
u $\underline{p}$ + (-ℓ)(1-p) = 0, hence p/(1-p) = ℓ/u. If $\underline{p}$ were smaller the
consistent person would not accept the bet, given ℓ and u.

As remarked in my Item 7 this presupposes, in terms of
modern decision theory, that utility is linear in the dollar
amounts -- an assumption rejected by Bayes' contemporary Daniel

Bernoulli and his Petersburg paradox. This difficulty, -- not
present in the more general approaches of Ramsey and Savage --
was recognized by their fellow-personalist DeFinetti in the
English-language revision (1964) of his pioneering "La
prévision..." (1937). It is perhaps true that money utility
is almost linear for small monetary gains and losses, and per-
sonal probabilities are revealed by the ratio l/u in a bet accepted
by a consistent (e.g., an appropriately experienced) bettor:
this is believed by both DeFinetti and Savage (1962). Others --
e.g., Tversky (1972) -- believe[3] that, typically, small monetary
differences fail to be discriminated by the subject. This re-
sults in faulty elicitation of the subject's probabilities by the
observer.

My second remark refers to the "objective" probability of
my Item 6. I would now prefer the word "intersubjective". When
a person considers a coin to be "symmetrical", or a sequence of
trials of an experimenter to be "repeated and independent";
when, in short, a person regards those random events as
"exchangeable" in DeFinetti's sense, this is revealed by the
(rational) person's choice: if two events are exchangeable he is
indifferent on which of them to bet. If two or more persons, all
rational, agree that those trials are exchangeable they will agree
(by definition) that certain prior probabilities are equal al-
though they may disagree about their size. Moreover: even in
the presence of this latter, prior, disagreement (provided only
that they agree which events have non-zero probability), their
agreement about exchangeability of successive trials will entail,
when these trials are sufficiently numerous, an almost complete

agreement on the size of the subjective posterior probability of each particular outcome of the experiment; (and also on the "as if" subjective probabilities of hypotheses, as mentioned above). For this posterior probability will be approached by that outcome's observed (hence "objective") relative frequency. To put it differently: each of the rational subjects whose prior choices reveal agreement that certain trials are exchangeable, will come closest to choosing the same bet a posteriori as the best, if each of them computes expected utility on the basis of observed relative frequencies (which are the same for all subjects) used as posterior probabilities.

Jacob Marschak

May 1974

University of California,
Los Angeles

UNIVERSITY OF CALIFORNIA

WESTERN MANAGEMENT SCIENCE INSTITUTE

INTERDISCIPLINARY COLLOQUIUM ON MATHEMATICS IN THE BEHAVIORAL SCIENCES

Los Angeles, February 11, 1966

Morris H. DeGroot

Stopping Rules

(Outline)

Consider an experimenter who can take independent observations $X_1, X_2, \ldots,$ sequentially from a population whose distribution involves some unknown parameters. After each observation X_n the experimenter can either stop sampling and receive a specified reward $r(X_1, \ldots, X_n)$ that depends on the values $X_1, \ldots, X_n$ that he has observed, or pay a specified price (typically a constant cost per observation) and observe X_{n+1}. It may also be true that at some stage (possibly random) the experimenter is forced to stop sampling and accept his reward. The total gain of the experimenter when he stops is the reward $r(X_1, \ldots X_2)$ that he receives minus the amount spent on sampling. The problem is to find a stopping rule that maximizes his expected total gain.

Distributional Assumptions and Notation

In the problems to be discussed, it is assumed that the observations $X_1, X_2, \ldots$ are normally distributed with unknown mean θ and variance 1. The parameter θ is assumed to have a normal prior distribution with mean μ and variance σ^2. Symbolically,

$$\theta \sim N(\mu, \sigma^2) .$$

The marginal distribution of an observation X is then $N(\mu, 1 + \sigma^2)$ and the posterior distribution of θ given $X = x$ is

$$\underline{N}\left(\frac{\mu + \sigma^2 x}{1 + \sigma^2}, \frac{\sigma^2}{1 + \sigma^2}\right).$$

The symbols φ and Φ denote the density function and the distribution function of the $N(0,1)$ distribution. The following function ψ occurs throughout the discussion:

$$\psi(t) = \int_t^\infty (x-t)d\Phi(t) = \varphi(t) - t[1-\Phi(t)].$$

It has the property that

$$\psi(-t) = \psi(t) + t$$

and its derivative is given by

$$\psi'(t) = -[1-\Phi(t)].$$

The inverse function ψ^{-1} is also used. For any numbers x and y, write

$$x \vee y = \{\text{the maximum of } x \text{ and } y\}.$$

Some Specific Problems

1. <u>Sampling without recall.</u> In this problem the experimenter's record when he stops is X_n, the value of the last observation that he has taken. There is a fixed cost c per observation. The experimenter's total gain when he stops is $X_n - nc$. His position at any stage of the sampling process is described by the triple (r, μ, σ^2) where r is the reward that he will receive if he stops without further sampling and (μ, σ^2) are the parameters of the current posterior distribution of θ.

Let $V(r, \mu, \sigma^2)$ denote the value to the experimenter of being in position (r, μ, σ^2). It is the expected total gain following the optimal procedure from position (r, μ, σ^2).

2. <u>Forced stopping</u>. In this problem there is no cost of sampling. However, at any stage of the process, if the experimenter decides to continue sampling, there is a fixed probability $p(0 < p < 1)$ that he will be forced to stop after the next observation and accept its value as his reward. His total gain when he stops is X_n. His position (r, μ, σ^2) at any stage and its value $V(r, \mu, \sigma^2)$ are as defined above.

3. <u>Sampling with recall</u>. If the experimenter stops sampling after having observed $X_1, \ldots, X_n$ his reward is $X_1 \vee X_2 \ldots \vee X_n$, the largest of the observations that he has taken. There is a fixed cost c per observation. The experimenter's total gain when he stops is

$$X_1 \vee \ldots \vee X_n - nc .$$

His position (r, μ, σ^2) at any stage and its value $V(r, \mu, \sigma^2)$ are as defined above.

4. <u>Variations on the above</u>. Further problems involve (i) discounted observations, (ii) guaranteed minimum rewards, (iii) a choice of populations from which to sample at each stage, (iv) a reward function of the form $X_n \vee X_{n-1} \vee \ldots \vee X_{n-k}$.

<div align="right">Carnegie Institute of Technology</div>

J. MARSCHAK

Do personal probabilities of probabilities have an operational meaning?

1. The operational meaning of personal probabilities is easily established when they are defined on the set of events that will influence the consequences of an ideally "consistent" ("rational") person's actions. That is, the personal probability $P(Z)$ of event Z for consistent Mr. Smith can be revealed (or at least approximated, as I shall explain presently) by his choices: not just by the purely linguistic exercise of asking him to name a number, as has been done by some contemporary experimenters, with results that can be used to predict future words, not future actions.

2. However, "Bayesian statisticians" use, in addition, personal ("prior") probabilities defined, not on the set of events, but on the set of probability distributions over this set. For example, DeGroot's μ and σ characterize a (normal) prior distribution of the parameter θ which, in turn, characterizes the distribution of the random sequence (x) of identically distributed variables, whose values do affect the decision-maker's payoff.

3. The operational meaning of a prior distribution of distributions, in terms of the decision-maker's choices, is not obvious. Yet, after recapitulating the meaning of personal (and also of objective) probabilities of events, I shall try to extend it to the prior distributions of distributions, much inspired by a brief exchange of opinions with our speaker today.

4. Let X be the set of states x of Nature (not controlled by the decision-maker). When x is in the subset Z of X we say that event Z has happened. Thus a probability measure P on X will determine the probabilities $P(Z)$, $P(Z')$, ... of the events $Z, Z', \ldots$. When a person takes an action whose consequence if Z happens, is preferable to its consequence if Z does not happen, we say that he bets on Z .

5. Now, a probability measure P on X is called personal with respect to Mr. Smith if, given any two events Z, Z', he prefers to bet on Z rather than on Z' whenever $P(Z) > P(Z')$, and is indifferent when ever $P(Z) = P(Z')$. [Note again: personal probability is revealed by what the person does, not by his stating verbally a number!] It has been shown by F. Ramsey (1926-28), B. De Finetti (1937), L.J. Savage (1954) that, for a person obeying the rules of logic supplemented by a few plausible consistency postulates, personal probabilities of events, in the sense just stated, do exist, along with a numerical utility function on the set of consequences of his actions. Utility is defined as a variable, whose expectations (computed on the basis of personal probabilities) the person's chosen action will maximize over the set of all available actions.

6. Moreover, in the case of certain ideally symmetrical, interchangeable events (ideal coins, ideally repeatable samples) rules of logic will make the personal probability of an event the same for all consistent persons, thus making this probability "objective".

7. T. Bayes himself (1963), followed by De Finetti [and also by R. Carnap (1962) in a paper presented in this Colloquium, March 10, 1961] considered personal probabilities as revealed by choices between bets involving

different monetary odds. Thus $P(Z) = p$ for a consistent Mr. Smith, if he accepts any bet on Z in which the ratio of his money losses to his money gains is smaller than $p/(1-p)$. But this presupposes, too narrowly, a linear utility function of money. Let me state a more general approach. First establish that Smith is indifferent between betting on a clock's twirled hand's stopping within any two equal arcs of the circumferance. This will reveal that he assigns certain ideal symmetry properties to the physical mechanism used. Now following the spirit of a suggestion by E. Borel (1924): if Smith prefers betting on event Z ("rain tomorrow") to betting on the hand's stopping within a 30° arc but not to betting on its stopping within a 60° arc, then for him

$$1/12 \leq P(Z) \leq 1/6 \ ,$$

— and so on in an obvious succession of steps. This "Borelian" procedure 5) is analogous to that of the ear-doctor's assessing your hearing capacity, or the analytical chemist's titration (and is subject to the same limitations except as an ideal).

8. Consider now n urns in which the proportions, p , of red balls are equal, respectively, to $p_1,\ldots,p_n$. Suppose you know that one and only one of the n urns is being used in a sequence of drawings. Let $f(p_i) = f_i$ be the personal probability you assign to the event ("hypothesis") that the urn used is the i-th urn. The function $f(\cdot)$ is your prior probability of the parameter p viewed as a n-valued random variable. We have

$$(1) \quad \sum_{i=1}^{n} f_i = 1 \ ; \quad f_i \equiv f(p_i) \geq 0, \ i = 1,\ldots,n.$$

You will never find which of the n events actually takes place, i.e., which urn is being used. Therefore the numbers f_i <u>cannot be established directly</u> by asking you to choose between bets on the i-th event and on the arcs of our dial.

9. But here is an indirect method. We shall draw balls, with replacement. Before the first drawing, I use on you the Borelian procedure to establish your personal probability that the ball will be red; call this number r_1. Then clearly the numbers f_i are related to known numbers $r_1, p_1, \ldots, p_n$ by the relation

(2) $$r_1 = \sum_{i=1}^{n} f_i p_i ,$$

subject to the constraints (1). Let $x_t = 1$ or 0 according as the t-th ball drawn is or is not red. Applying the Bayes theorem (an identity implied by the definition of conditional probabilities), the n posterior probabilities, given the result x_1 of the first drawing, are (summing here and henceforth over i)

$$\left.\begin{array}{ll} f_i p_i / \Sigma\, f_i p_i & \text{if } x_1 = 1 \\ f_i(1-p_i)/\Sigma\, f_i(1-p_i) & \text{if } x_1 = 0 \end{array}\right\} \quad i = 1,\ldots,n$$

Your personal probability, after the first drawing, that the second drawing will be red ($x_2=1$) is therefore equal to

$$r_2 = r_2(1) = \Sigma\, f_i p_i^2 / \Sigma f_i p_i \qquad \text{if } x_1 = 1$$

$$r_2 = r_2(0) = \Sigma\, f_i p_i(1-p_i)/\Sigma f_i(1-p_i) \quad \text{if } x_1 = 0.$$

I know x_1; and I establish your r_2 by the Borelian procedure; we have now added one more equation to (1) and (2), with the f_i as unknown and $r_1, r_2, p_1, \ldots, p_n$ known. In general, writing

$$y_t = x_1 + x_2 + \ldots + x_t , \quad t=1,2,\ldots$$

$$y_0 = 0,$$

your personal probability (ascertainable by the Borelian procedure),
r_{t+1}, that the $(t+1)$-th ball will be red is

(3) $r_{t+1} = r_{t+1}(y_t) =$

$$= \frac{\Sigma \, f_i p_i^{y_t+1} (1-p_i)^{t-y_t}}{\Sigma \, f_i p_i^{y_t} (1-p_i)^{t-y_t}} \qquad , \quad t=0,1,\ldots$$

If $t+1= n-1$, i.e. if the Borelian procedure was applied $n-1$ times
(i.e., $n-2$ drawings were made) we have as many equations as we need to
determine $f_1,\ldots,f_n$, subject to (1). The numbers r_{t+1} elicited from
the consistent man **must** also obey the **inequality** in (1). Moreover, his
values for $r_n, r_{n+1}, \ldots$ must be such that the system of equations and
the inequality are satisfied by some values $f_1,\ldots,f_n$. Thus if $n=2$ and
$p_1 \neq p_2$ then (without loss of generality) $p_1 > r_1 > p_2$ and

$$f_1 = (r_1-p_2)/(p_1-p_2)$$
$$f_2 = (p_1-r_1)/(p_1-p_2) \, ,$$

and hence, for $n=2$, all the subsequent values $r_2, r_3, \ldots$ are fixed without
ever making a drawing - provided you are "consistent".

However, let p range continuously from 0 to 1. Then the vector
$[f_i]$ is replaced by a density function $f(p)$; or, more generally, we try
to find the distribution function $F(p)$ from the functional equations

(3') $$r_{t+1} = \frac{\int_0^1 p^{y_t+1} (1-p)^{t-y_t} \, dF(p)}{\int_0^1 p^{y_t}(1-p)^{t-y_t} \, dF(p)} \qquad , \quad t=0,1,\ldots,T$$

Can we conjecture that, with T finite, F cannot be ascertained from the sequence $r_1, \ldots, t_{T+1}$ exactly? And that an approximate solution will, in some sense, converge to F as T increases?

10. Now return to the case of a finite number n of hypotheses (i.e. of values of the parameter p). We can modify our procedure so that no actual drawings need take place, even when n > 2. I can use the Borelian dial to establish the personal probability $q(t,y_t)$ which you assign to the following event: in a sequence of t drawings without replacement, y_t balls will be red. Then

$$(4) \quad q(t,y_t) = \sum_{i=1}^{n} f_i p_i^{y_t} (1-p_i)^{t-y_t} \binom{t}{y_t} \, ;$$

here the integers t and y_t can vary arbitrarily, independently of n, provided $t \geq 1$ and $0 \leq y_t \leq t$. We can therefore produce as many as n equations of the form (4), all linearly independent in the f_i. If you are consistent, the f_i will satisfy the constraints (1). Moreover any additional equation of the form (4) elicited by another application of the Borel dial should be satisfied by the same values of the f_i.

11. Again, similarly to the case of (3'), no finite number of such applications will suffice to determine the distribution function F(p) when p is continuous and (4) becomes

$$(4') \quad \frac{q(t,y_t) \cdot y_t!(t-y_t)!}{t!} = \int_0^1 p^{y_t} (1-p)^{t-y_t} \, dF(p) \, ,$$

with a known number on the left side and a sort of weighted Beta-function (with unknown weights) on the right side. Again: will successive

approximations converge to F as the number of pairs (t, y_t) increases?

12. All this was intended, not to suggest psychological experiments but merely to show that operational meaning <u>can</u> be attached to prior probabilities of statistical parameters. Psychological experiments -- observing $q(t, y_t)$ in (4) -- will most probably yield the result that no mortal is consistent. Except, of course, the ideal statistician.

13. In a subsequent discussion, Professor Glenn Graves raised the following subtle question. The original consistency postulates have been defined for events on which <u>direct</u> bets were possible. The prior and posterior probabilities of the Bayes theorem had to be understood accordingly. On the other hand in stating that the consistent man's responses r_t, y_t must satisfy certain constraints (as in the paragraphs following equation (3) and equation (4), respectively), Bayes theorem was applied to "events" (viz., sets of values of the parameter p) on which only <u>indirect</u> bets can be taken, in the sense described above. But: do those constraints follow from the original postulates, or must the latter be replaced by stronger ones?

A Circular Letter

I would appreciate it if you would give your opinion on the question raised in the enclosed mimeographed note of mine. It was written in connection with a talk delivered by Morris DeGroot in which he used the Bayesian approach. The note is only loosely related to the special topic of that talk (Stopping Rules). It deals with a general difficulty which has bothered me and, in the last paragraph, with a further logical question raised by G. Graves.

If you would care to comment, kindly state in the body of your letter or on the enclosed paper, whether you would permit me to circulate your comments to these same people.

9 May 1966   University of California,

Los Angeles

KARL BORCH

The mathematics of your intriguing note about personal probabilities is obviously related to what the statisticians used to call "The Problem of Moments" some 30 years ago.

Your equations (3') make it possible to determine the $T + 1$ first moments of the distribution $F(p)$ -- or if you like, the $T + 1$ first coefficients in the power series of the characteristic function. The other coefficients can be chosen arbitrarily. Hence there will be an infinity of distributions $F(p)$ compatible with a consistent sequence $r_1 \ldots r_{T+1}$.

The convergence should follow from the fact that a distribution is uniquely determined by its moments -- or if you prefer -- by its characteristic function.

These superficial remarks do of course gloss over a number of difficulties. The moments determined by (3') may, for instance, give a distribution defined over a greater domain than $(0,1)$. I ı not certain how this should be interpreted.

You should, of course, feel free to circulate this letter if you think it is useful.

6 June 1966   The Norwegian School of Economics
                                             and Business Administration, Bergen

HERMAN CHERNOFF

The question you raised in the colloquium comes up also in problems of Empirical Bayes or Compound Decision Theory. There is a basic Identifiability problem underlying it. In the binomial case one can estimate

F(p) by knowing its moments. Thus in equation (4'), the sequence

$$q(t,t) = \int_0^1 p^t \, dF(p) \qquad t = 1,2,\ldots$$

determines F, and a finite subsequence can be used to approximate F.

In Compound Decision problems it is typical that one is mainly interested in a particular function of F, e.g., $\int p \, dF(p)$. Suppose as a typical problem, that n coins arrive each with a possibly different value for the probability of head, p, (presumed to be independently selected from an unknown distribution F(p)). Suppose each coin is tossed once. Then the proportion of heads observed is an estimate of $\int p \, dF(p)$.

If the appropriate decisions depended on $\int p^2 dF(p)$, the experimental setup described above seems inadequate to get a "good" estimate of the desired quantity.

Another example of lack of identifiability which is near and dear to me stems from a problem in scoring multiple choice questionnaires. Suppose a question can be answered True or False. Assume that each student answers the question correctly if he knows the answer and guesses at random otherwise. If 50% of the students answer incorrectly, it is evident that almost no one knew the answer and an appropriate procedure would be to mark everyone wrong even if they achieved the correct answer. The appropriate way of handling the students depends on your estimate of λ, the proportion of the students who know the correct answer; and λ may be estimated in terms of the directly observable proportion of students who answer correctly. A more sophisticated version gives lack of identi-fiability. Suppose that there are three choices A, B, C of which A is

correct. The model assumes that students are of four types. Those who know the answer and answer correctly (proportion λ_1). Those who know the answer is A or B and guess with probability 1/2, 1/2 between them (proportion λ_2), those who know the answer is A or C and guess with probability 1/2, 1/2 between them (proportion λ_3). Finally there are those with no knowledge and guess with equal probability among A, B, C (proportion $\lambda_4 = 1-\lambda_1-\lambda_2-\lambda_3$). If λ_1, λ_2, λ_3, λ_4 were known, a "good" way to grade the individual students would be easily determined. The data only supply enough information to estimate two of the three unknown parameters. The multiple choice scorer can resolve his question partially by using a minimax approach. However, for the Bayesian to recover the entire prior distribution, it is required that a sufficient body of experimental data be available to be tapped.

3 June 1966   Stanford University

MORRIS DE GROOT

Professor Marschak's comments are very closely related to the important work of de Finetti (1937), revised and translated in Kyburg and Smokler (1964), and discussed also by Savage (1954), Ch. 3, Sec. 7. This work, which gives strong support to the Bayesian theory of statistics, shows that if a person's probabilities on the _outcomes_ of a sequence of coin tossings satisfy certain conditions of symmetry, or exchangeability, then his probabilities can be represented by the conditional distributions given a fictional "unknown p" of the coin, together with a unique "prior distribution of p." I agree with

Professor Marschak in emphasizing that, in general, an infinite sequence of tosses is required to learn this prior distribution.

28 February 1966   Carnegie Institute of Technology
                                    and Stanford University

ROBERT DORFMAN

Thank you for sending me your note on personal probabilities of probabilities. I found it very ingenious and very convincing, and about as lucid a description of how to generate subjective probabilities as I have yet seen.

My only slight divergence from you is my conviction that the world about which we make decisions consists of nothing but probability distributions. If I bet indifferently on red and black at roulette, I am disclosing my belief that the wheel is well-balanced, whereas my neighbor who bets a system discloses his belief that the wheel obeys a nonstationary stochastic process. If I seem to bet on "rain tomorrow" I am really expressing my belief that today's weather system is such that probability of rain tomorrow is high. In this sense all we ever reveal is our personal probabilities of probabilities, so that the case you discuss is the fundamental case that people have been discussing (mostly implicitly) all along. What the recorded experiments disclose is actually your r_1 rather than p_1.

No objection to circulation if you think it worthwhile. My personal probability of the probability of that event is low.

19 May 1966   Harvard University

WARD EDWARDS

I have read with much interest your comments on DeGroot's paper, and have a few comments. Please feel free to make any use of them that may be convenient.

I am in no serious disagreement with your conclusion, or with the arguments that lead you to it. However, I feel that you are working unnecessarily hard by defining an unnecessarily restricted universe to be interested in. You have chosen to talk only about questions in which the population characteristic of interest is now knowable. Many questions are like that, at least in practice. But many others are not at all like that, and the situations in which you can conveniently get at the population parameter offer useful conceptual devices for thinking about those in which you cannot.

The problem arises first in your second paragraph, where you talk of prior distributions as sets of distributions over sets of events. When I think of a prior, or a posterior, distribution, I think of it as my opinion about some well-defined issue about which I am uncertain. If, for example, I am uncertain about whether this bookbag contains 700 red and 300 blue chips, or 700 blue and 300 red, my uncertainty should be describable by a number. (It is almost meaningless to call this number a prior or a posterior distribution or opinion. All opinions are posterior to some information and prior to other information. Prior distributions occupy no special status that I know of in Bayesian thinking.)

My opinions can be modified by means of information that for me bears on them. Sometimes that information bears on them so potently

(in my opinion) that my posterior distribution would approach 1 (for
discrete hypotheses) or would have a peak higher than any preassigned
number (for continuous hypotheses). In the example, such an item of
information could be obtained by dumping out the bag and counting the
chips. Other kinds of information, such as might be obtained by sampling
with replacement, are less convincing. I see no sharp formal lines
differentiating overwhelming from non-overwhelming evidence. Of course,
such lines can easily be constructed, and it makes rather little differ-
ence what choice of operational definition of "overwhelming" is used,
so long as the arbitrariness of that definition is recognized. For
illustration, I shall arbitrarily define evidence overwhelming for a
discrimination between two hypotheses as evidence sufficient to change
prior odds of 1:1 into posterior odds of at least 1,000,000:1 as between
that pair of hypotheses. For me, counting the chips in the bookbag easily
meets that test -- given the truth of the model of the data-generating
process that I am tentatively working within.

Opinions characterize me, not the bookbag. My opinions about either
the proportion of red chips in the bookbag or the probability that the
next chip to be sampled will be red can be defined, and measured, only
by observing my behavior. (Discussion of what behavior to observe comes
later in this letter.) I see absolutely no formal difference between my
opinions about the population parameter and about the identity of the
next sample. There is, of course, an important practical point: I have
a formal model that implies for each possible bookbag composition, what
the probability is that the next sample will be red. I hold that model

as a working hypothesis with high probability. Thus some internal consistency rules link opinions about samples (data) with opinions about bookbag compositions (hypotheses). However, these internal consistency rules work in both directions; specification of $P(D|H)$ by itself specifies neither $P(H)$ nor $P(D)$, but to some extent constrains both, given the existence of some data.

Now we come to the question of what observations you might make in order to discover my opinions. Such observations are numerous. By far the simplest procedure, of course, is simply to ask me. For some reason you object to this. You make a distinction that I cannot understand between what you call words and what you call actions. Why aren't words actions? You seem to feel that if I discover that a subject is indifferent between betting on A and betting on $\overline{A}$, then I am justified in saying that A and $\overline{A}$ are equally likely, for him. Yet you deny him the privilege of making, or at any rate of communicating to you, the same inference.

I feel that many different actions, some performed with tongue and some not, some with immediate consequences and some not, are suitable for indicating what my, or anyone else's opinions are. If the person being studied is an ideally consistent man, then those opinions will obey all appropriate consistency rules, including the ones that permit specification of coherence between words and betting behavior. If he is a real man, he will of course be inconsistent. It is an empirical question, not to be answered from the armchair, whether that subset of his behaviors defined by linguistic responses having no immediate consequences will or will not be consistent with other subsets of his behavior, such as are

studied in so-called choice experiments. I know of absolutely no empiri-
cal evidence indicating that such verbal responses are less consistent
with choices than choices are with one another; in fact, if anything I
think the evidence is in the other direction.

The set of responses you have proposed is, of course, in principle
a usable set. Of course, it doesn't really meet your objection to the
interpretation in terms of bets without the auxiliary Borel dial. You
require, as a first step, establishment of indifference between equal
arcs. Indifference is not observable in human behavior; it must be
inferred from easily switchable preferences, as in the Davidson, Suppes,
and Siegel experiment. But given that kind of procedure and a consistent
subject, nothing more is required to obtain utility-of-money functions.
And given those functions, the definition of odds in terms of acceptable
bets works fine, using utiles instead of dollars in the relevant ratios.
In other words, I see no reason for constant reference to the Borel dial.
Instead, the indifference-defining operation can be used to measure the
utility of money, and thereafter utilities instead of dollars can be used
in a bet-based definition of probability. Or, instead, you could simply
ask consistent people to estimate numbers... Of course any of these pro-
cedures, applied to real men, will produce inconsistencies. Any single
procedure will produce internal inconsistencies; any pair of procedures
will produce inconsistencies between the results of the two procedures.
From this point of view, I hold no one procedure inherently more valid
than any other. Your procedure of using sample urns, observing prior
and posterior distributions, and inferring what the "first" prior must

have been, is feasible. But why be so roundabout? Why deny yourself the freedom to dump out those urns and count the balls in them? Then you can indeed observe what the true event is, and so can settle bets. I see no difference between that procedure and the sampling procedure you propose, except that the sampling procedure is a lot more complex. But, of course, I see no difference of formal or philosophical status between the events that I have been calling data and those that I have been calling hypotheses. And that brings us back to where we started.

I know nothing about Professor Graves's question.

18 May 1966   University of Michigan

T.S. FERGUSON

Here are the comments you requested on your remarks on Probabilities of Distributions.

1. Let $X_1, X_2, \ldots$ be independent Bernoulli trials with probability θ of success, and let $\tau(\theta)$ be a prior distribution. The mathematical problem may be stated: given sufficiently accurate information on the marginal distribution of $X_1, X_2, \ldots, X_n$, can one determine $\tau(\theta)$ sufficiently accurately? For fixed n, obviously not: for n = 1, all that can be determined accurately is $P(X = 1) = \int \theta d\tau(\theta)$, namely, the mean of τ. However, from the marginal distribution of $X_1, \ldots, X_n$, one can determine the first n moments of τ, so that for sufficiently large n and sufficiently accurate information on the first n moments, τ may be determined as accurately as desired, because it, being a bounded distribution, is determined by its moments.

2. A procedure for estimating τ is as follows. Let $X_n = \frac{1}{n} \sum_1^n X_i$. Then,

$$P(\bar{X}_n \le t) = \int_0^1 P_\theta(\bar{X}_n \le t) \, d\tau(\theta)$$

and

$$P(\bar{X}_n \le t) \to \begin{cases} 0 & \text{if } t < \theta \\ 1/2 & \text{if } t = \theta \\ 1 & \text{if } t > \theta \end{cases}.$$

Hence

$$P(\bar{X}_n \le t) \to \frac{\tau(t^+) + \tau(t^-)}{2} \underset{=}{\overset{\text{def}}{=}} \tau(t)$$

so that if $\underline{n}$ is chosen so large that $\left| P(X_n \le t) - \tau(\underline{t}) \right| < \varepsilon/2$, and $P(X_n \le t)$ is determined to within $\varepsilon/2$, then $\tau(t)$ is determined to within ε, as the quantity $P(X_n \le t)$.

3. This problem is related to the problem of identifying and estimating a mixing distribution. See for example Teicher [Ann. Math. Stat., 1963, pp. 1265-1269]. The general problem is: given $F_\theta(x)$, to estimate $\tau(\theta)$ from a sample from the distribution $H(x) = \int F_\theta(x) d\tau(\theta)$. This corresponds to the above problem for n = 1. The distribution τ can be so estimated for many kernels, $F_\theta(x)$. See Gaffey [Ann. Math. Stat., 1959, pp. 198-205] for the case where $F_\theta(x)$ is normal with mean θ and variance 1. A student of mine, Carl Maltz, has found corresponding methods for other families of distributions -- to appear in his Ph.D. thesis.

4. The distribution with density

$$f(x|\theta) = \begin{cases} \frac{1}{2}(1 + \theta x) & \text{for } -1 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

where $-1 \leq \theta \leq 1$, like the Bernoulli, does not lend itself to estimating

the mixing distribution, even though there are an infinity of values,

$h(x) = \int f(x|\theta)d\tau(\theta)$ to be used. This is because $h(x) = \frac{1}{2}(1 + x\int\theta d\tau(\theta))$,

so again only the mean of τ may be determined. This corresponds to

$n = 1$.

If n is allowed to be arbitrary, then again τ may be estimated.

One procedure for accomplishing this is to transform to the Bernoulli

case, calling $X \geq 0$ a success, and $X < 0$ a failure. Then since the dis-

tribution of the probability of success, $\frac{1}{2} + \frac{\theta}{4}$, may be approximated, so

may the distribution of θ.

This procedure extends to all cases where θ is identifiable from

$F_\theta(x)$ for x in some denumerable set, D (i.e., when θ can be found knowing

the numbers $F_\theta(x)$ for x in D). For example, if the sample space is

Euclidean, and if θ is identifiable, then τ may be approximated.

15 June 1966   University of California,
                                            Los Angeles

KOICHI MIYASAWA

If I am not misunderstanding Professor Marschak's note, his issue

comes from the following postulate: in order that the personal prob-

ability of an event Z has an operational meaning, the event should be a

real one. Here by a real event I mean one about which the person can

know it obtains or not after all.

If we admit non-real events in determining their personal probabili-

ties by the choice behavior of the person among bets on these events,

then it seems to me that there is no need to raise a question on the operational meaning of personal probabilities of probabilities.

I would like to admit that the reasonable person Mr. Smith can contemplate an imaginary or non-real bet which is assumed to give a prize to him if the true ratio p of the red balls in the urn lies between, say, 10% and 30% even though he might never know the true value of p and can show his choice between the bet and a bet which is defined by means of a clock's twirled hand. If we do not admit such bets on non-real events in operational definition of personal probability, I do not think the personal probability, with respect to Mr. Smith who is going to board the plane, of a crash of the plane can have an operational meaning, since Mr. Smith cannot realize the event Z that the plane will crash obtains or not until he safely arrives at the destination by the plane or dies by a crash of the plane.

21 May 1966   University of Tokyo

PAUL RANDOLPH

... Although I have no answers to the questions posed by Marschak, I do have a few comments to make and a few additional questions to pose. In his comments Marschak suggests that the prior distribution function be denoted by F. It has been suggested by many (e.g., see Good, 1965) that F be a beta form, proportional to $p^{m_1}(1-p)^{m_2}$ where $m_1 > -1$ and $m_2 > -1$. If this assumption be made, can the two parameters, m_1 and m_2, of F be determined from r_{t+1} using equation 3' of Marschak or from equation 4'? I think not. For example, suppose t = 0. Then since

$$\int_0^1 p^{m_1}(1 - p)^{m_2} \, dp = \frac{m_1! \; m_2!}{(m_1 + m_2 + 1)!}$$

we have the probability that the first ball is red as

$$r_1 = \frac{m_1 + 1}{m_1 + m_2 + 2}$$

and, of course, the probability of not red is

$$\bar{r}_1 = \frac{m_1 + 1}{m_1 + m_2 + 2} \; .$$

Can one specify r_1 and $\bar{r}_1$ and from this determine m_1 and m_2? The answer is obviously no, and thus the prior distribution function F is still unknown.

Of course, if a red is observed on the first sample, the probability of a red the second time is now

$$r_2(1) = \frac{m_1 + 2}{m_1 + m_2 + 3}$$

and the probability of a not red is

$$\bar{r}_2(1) = \frac{m_2 + 1}{m_1 + m_2 + 3} \; .$$

If now, the experimenter had been required to state his personal probabilities of a red on the first draw and the conditional probability of a red on the second draw, given that a red had already been observed on the first draw, then we have m_1 and m_2 and thus F.

However, I think this is putting quite a strain on the imagination of the experimenter. Perhaps there is another way out.

If we examine

$$r_{t+1}(y_t) = \frac{m_1 + y_t + 1}{m + t + 2}$$

$$r_{t+1}(y_t) = \frac{m_2 + (t-y_t) + 1}{m + t + 2}$$

where $m = m_1 + m_2$, we see that m_1 and m_2 play a role similar to that of y_t and $t-y_t$, respectively. Can we conclude then that m_1 and m_2 represent the number of reds and non-reds that the experimenter thinks he would have seen from a sample of size m? Or, since

$$r_1 = \frac{m_1 + 1}{m + 2} \quad ,$$

if the experimenter is willing to specify his personal probability, r_1, can we also let him specify m as his degree of confidence that this is the correct r? If so, would $m = 0$ imply no confidence? For $m = 0$, m_1 and m_2 can vary from -1 to +1. It is a bit awkward to try to interpret a value such as $m_1 = .5$ to be the number of red balls one can expect in 0 draws. Also, what is the interpretation of negative m?

If this is difficult, suppose the experimenter is unwilling or unable to guess r but is willing to try to specify a prior distribution for r, say $\phi(r)$. If this is done, will this be enough to specify F?

Do you suppose the following is fair: Suppose the experimenter has no knowledge and no confidence. Thus, he may take $r_1 = 1/2$ and $m = 0$. This implies that $m_1 = 0$ and $m_2 = 0$ and, therefore, the prior density is the uniform, that is

$$f(p) = \begin{cases} 2 & p \in [0,1] \\ 0 & \text{otherwise .} \end{cases}$$

Suppose the first observation was red. Then we would expect

$$r_2(1) = 2/3 .$$

But, suppose that after seeing the first observation experimenter says "oops, I think I goofed. I think r_1 should be something else, such as $r_1 = .60$. Furthermore since I have some information from one observation I will take $m = 1$." If he is allowed to do this, then we would allow him to change m_1 and m_2 to

$$m_1 = .8 \qquad m_2 = .2$$

this makes

$$f(p) = \frac{2!}{(.8)!(.2)!} p^{.8}(1 - p)^{.2}$$

which in turn gives

$$r_2(1) = \frac{2.8}{4} = .7 .$$

Continuing, after observing the second observation should we allow him to change his mind again regarding r_1?

A similarly interesting set of questions arises when urns do not contain balls but disks with numbers, such that $p = (p_1, p_2, \ldots, p_k)$ is the vector of probabilities p_j that a disk drawn at random is equal to j, $j = 1, 2, \ldots, k$. It is evident that p is an element of the simplex

$$S = \{p : p_j \geq 0, \ \sum_{j=1}^{k} p_j = 1\} .$$

Then, suppose n observations are made, $x_1, x_2, \ldots, x_n$, of which n_j are equal to j; that is, n_j is the cardinality of the set $\{x_i : x_i = j\}$. Then, the probability that the $(n+1)^{th}$ ball is j is

$$r_{n+1}(j) = \frac{\int_S p_1^{n_1} p_2^{n_2} \cdots p_j^{n_j+1} \cdots p_k^{n_k} \, dF(p)}{\int_S p_1^{n_1} p_2^{n_2} \cdots p_j^{n_j} \cdots p_k^{n_k} \, dF(p)} \quad .$$

Suppose as before we let f(p) be proportional to

$$p_1^{m_1} p_2^{m_2} \cdots p_k^{m_k}$$

where $m_j > -1$ for all j. Then by the Dirichlet integral (Dirichlet, Comp. Rend. Acad. Sci., 1839) we have

$$r_{n+1}(j) = \frac{m_j + n_j + 1}{k + m + n} , \qquad j = 1, 2, \ldots, k$$

where $m = \sum_{j=1}^{k} m_j$. Then,

$$r_0(j) = \frac{m_j + 1}{k + m} \quad .$$

We could give values to $r_0(j)$, $j = 1, 2, \ldots, k$, but as before, this would not give F(p). If we could specify the personal probabilities $r_0(j)$ and also the confidence value m, then of course we know F.

Suppose our personal feelings are only that $r_0(j)$ can be approximated quite nicely by a normal distribution, that is,

$$r_0(j) \approx \int_{j-1/2}^{j+1/2} \frac{1}{\sqrt{2\pi} \, \sigma} e^{\frac{-(x - \mu)^2}{2\sigma^2}} \quad .$$

Indeed, this is suggested by J. Heller (1960) for the scheduling problem. If μ and σ^2 are not known, can we take n observations, find the sample mean x_n and variance s_n as the maximum likelihood estimates of μ and σ^2, respectively, and use these for computing good values of $r_0(j)$? Perhaps, this is feasible when one wishes to use these prior distributions for tests of hypotheses, such as stopping rules.

On the other hand, would we want to digress one step and require the experimenter to specify a prior distribution for μ and σ? What kind of an operational meaning could we ascribe to these probabilities? I do not know.

One last remark: I have developed stopping rules for the multinomial. For the case of complete ignorance ($r_j = 1/k$, $j = 1,\ldots,k$) and no confidence ($m = 0$) I found that the convergence was exasperatingly slow. I certainly cannot show that it is right. I am merely using the age old engineering motto "if it works, it is right." So far, the results are very encouraging.

26 September 1966   New Mexico State University

LEONARD J. SAVAGE

Almost all that you inquire about has long been well studied under the rubric of exchangeable processes. A now obsolete term is sequences of equivalent events. Not the earliest, but one of the most thorough and important references, is de Finetti's masterpiece, "La Prévision" (1937), in the Institut Henri Poincaré, which was translated and brought up to date under de Finetti's supervision in the anthology edited by Kyburg and

Smokler (1964). Section 3.7 of my book, <u>The Foundations of Statistics</u> (1954)
is devoted to the same topic. Though these references should put you
in almost complete possession of the facts, and though others to whom
you circulated your manuscript, "Do personal probabilities of probabili-
ties have an operational meaning," may already have covered the ground
pretty well, I shall make some of the obvious applications of the theory
of exchangeable sequences of events to various passages of your manuscript.

1. With reference to your Item 2, many Bayesian statisticians,
especially those directly influenced by de Finetti, recognize clearly
that references to unknown θ may, depending on the context, be merely
a figurative way of describing a sequence of dependent random variables,
or something of the sort.

If, for example, I say that the x_i are normally distributed with
unit variance around θ and that θ is for me normally distributed around
μ with standard deviation σ, I may mean only that the x's are distributed
as they would be were there an actual physical constant θ about which my
opinion was as described, and with knowledge of which, the x_i's would for
me have independent normal unit distributions about θ. The actual upshot
of this is that the x_i's for me are variables with a joint normal dis-
tribution such that each has mean μ and variance $1 + \sigma^2$, and such that
the covariance between pairs of the x_i's is σ^2.

2. Your Item 3. Some of us Bayesians believe that there is no
intellectual need to, or possibility of, introducing any other kind of
probability than personal probability. That is not a thesis to be argued
here; for the moment I want only to point out that the thesis cannot

even be heard, let alone judged, without an understanding of exchange-
able events.

To illustrate, for us, the situation alluded to in your Item 6 is
this. In certain cases, a person will judge a sequence of events to be,
for him, both exchangeable and independent with a specified probability.
These assumptions together with the rules of logic imply all the prob-
abilities associated with this sequence of events. These probabilities
are "objective" in the trivial sense that any person who shares the
original person's underlying opinions will share with him the implica-
tions of those opinions -- that can be said about any system of opinions.

3. Your Item 7. Historical questions can be delicate and dangerous.
I have read Bayes' paper but discerned no evidence that Bayes regarded
probability as personal, or subjective. It would be worth some trouble
to document the point one way or the other.

4. Your Item 9. The indirect method initiated here is a little
more complicated and confusing than need be. It would be enough to ask
the person once and for all for his personal probability that all k of
the first k balls drawn will be red, for each k from 1 through n-1.
Calling these numbers $r(k)$, the numbers f_i are related to the known
numbers $r(k)$, $p_1, \ldots, p_n$ by the relations

$$(2') \qquad r(k) = \sum_{i=1}^{n} f_i p_i^k , \qquad k = 0, \ldots, n-1$$

where it is to be understood that the heretofore undefined $r(0)$ is 1.
If none of the p_i are equal to each other, this system of n linear
equations and unknowns had a unique solution, according to the theory

of the van der Monde determinant. Just when it has a nonnegative solu-
tion, I do not know but could perhaps look up.

A deeper analysis, more in the spirit of your (3') is that the
infinite sequence of r(k) constitute the moments of F, which according
to Hausdorff are sufficient to characterize F.

Such quantities as those in the numerator and in the denominator
of (3') are obviously inferable from moments, and of course have the
moments as special cases. Not just any sequence of numbers r(k) can
be the moments of a distribution; your (1) corresponds to the monotony
of F, which must be respected. It has been known at least since Hausdorff
that the necessary and sufficient condition for this is that the quanti-
ties that appear in (3') all be nonnegative. This is commonly expressed
in two ways as follows:

$$0 \le E(p^a(1 - p)^b) = E(\Sigma_c \binom{b}{c} (-1)^c p^{a+c})$$

$$= \Sigma_c (-1)^c \binom{b}{c} r(a+c) \equiv Q(a,b) .$$

But $Q(a,b)$ is the probability of a reds followed by b blacks. Whence
$Q(a,b) = Q(a+1,b) + Q(a,b+1)$. Therefore,

$$Q(a+1,b) = Q(a,b) - Q(a,b+1) \equiv -\Delta_b Q(a,b)$$

$$Q(0,b) = r(b) \qquad \ge 0$$

$$Q(1,b) = -\Delta_b r(b) \qquad \ge 0$$

$$Q(2,b) = \Delta_b^2 r(b) \qquad \ge 0$$

$$. . . .$$

$$Q(a,b) = (-1)^a \Delta_b^a r(b) \ge 0 .$$

5.   End of your Item 9. As has already been implied, the entire
sequence of r(k) is sufficient to determine F. Your r_j are a random
sequence, but the first T elements of it do always exactly determine
the first T moments r(k). Therefore the entire sequence of r_j, exactly
like the r(k), do determine the F, and it is actually true that less
and less latitude for F is available as T increases (see Shohat and
Tamarkin, The Problem of Moments, p. 77 ff., Amer. Math. Society, New
York, 1950). Occasionally, a finite number of r(k), and therefore of
r_j, is sufficient to determine F exactly. If, for example, $r(2) = r(1)^2$,
then necessarily F is entirely concentrated at the point r(1), so r(k) =
$r(1)^k$ for all k. Again, if the person feels certain that the balls are
either all balck or all red, that will be promptly revealed by the con-
dition that r(1) = r(2). I think of a few other exceptional cases by
combining the ones already mentioned. The general situation can probably
be dug out of the book by Shohat and Tamarkin.

6.   Your Items 10 and 11. The probability that a specific number
y_t of balls will be red in the first t drawings is, except for an uninter-
esting binomial-coefficient factor, the same as the probability that a
specific subset of balls with y_t numbers will consititute exactly the
red ones. These latter numbers, my Q(a,b), are for some purposes easier
to deal with, because the binomial coefficient is left out. The question
raised in your Item 11 is now seen to be the same as the one discussed
in my preceding point.

7.   It is important to realize that the question with which your
note ends is to be answered in the negative. No new postulates are needed.

If a person associates probabilities with a sequence of events x_i, and

if the events are exchangeable, that is if the probability for the person

of any finite pattern of successes and failures depends only on the

numbers of successes and failures involved and not on their order, then

(and only then) will there be an underlying F.   The entire system of

inequalities needed is only that the $Q(a,b)$ be nonnegative, but $Q(a,b)$

is simply the probability for the person of a specific sequence involving

a successes and b failures.

15 June 1966   Yale University

ROBERT SCHLAIFER

'If we are talking about subjective probabilities as a part of a

methodology for thinking through to decisions and not as part of a model

for predicting decisions, then I am afraid that my reaction to your dis-

cussion of Mr. de Groot's talk is that I just cannot see or feel that

there is any real problem to be discussed.

First, it seems to me clear that many probabilities for observable

events cannot be verified by actually observing betting behavior.   If I

think that promotional campaign A gives a 1/2 chance of "success" while

promotional campaign B gives a 3/4 chance of success, and if after taking

costs into account I analyze my decision problem and decide to use strategy

A, you cannot observationally verify the subjective probability 3/4 that

I assigned to success with promotional campaign B.   But when I am analyz-

ing my decision problem I feel that both my probabilities have exactly

the same kind of meaning to me; and since I am making the decision, I

don't really care at all whether some observer will or will not be able to "verify" one, both, or neither of my probabilities after my decision has been made.

As regards distributions of unknowable parameters of processes such as the Bernoulli, I am again unable to feel any real problem for a decision maker as opposed to a decision observer. I can imagine betting on the event r successes in n trials just as well when I know that the n trials will not really be made as I can when I know that they will really be made -- the only thing that counts is the fact that they have not yet been made -- the meaning to me of my probability for r heads in n trials does not depend at all on whether the trials will actually be made. I'm not quite so sure that I know what I mean by a fraction p of heads in an infinite number of trials, but I can always think about the fraction p' in a googolplex of trials and then argue purely mathematically that I will make no error of practical interest if I assign to the parameter p the same distribution that I assessed for p'.

You are more than welcome to make any use you like of these remarks, for the triviality of which I apologize.

27 June 1966   Harvard University

ROBERT L. WINKLER

Please excuse my delay in responding to your "Round Robin." It seems to me that the verbal approach need not be rejected, although it apparently has been rejected in the development of the personalistic theory (de Finetti, 1937; Savage, 1954). The criticism of the verbal

approach, as I see it, centers around the claim that we have no way of knowing if the verbal answers are in accordance with the assessor's beliefs and judgments -- indeed, the assessor does not necessarily have any incentive to make his answers correspond with his beliefs and judgments.

In light of recent work by de Finetti regarding scoring rules, or penalty functions, it appears that personal probabilities _can_ be given an operational meaning in terms of verbal answers. If a reward (or punishment) is determined from the assessor's answers according to some scoring rule which is so constructed as to oblige the assessor to make his answers be in accordance with his beliefs and judgments, then the element of incentive is present and the criticism removed.

Unfortunately, with the penalty functions, as with the betting rules, knowledge of an "actual value" is necessary to implement the methods. The "actual value" is needed to determine the winner of the bet or to determine the score obtained (and the resulting reward or punishment) through the penalty functions. Unless other penalty functions (e.g., dependent upon sample results) can be developed, then, we are still faced with the problems discussed in your note.

A final note is that the problem of "actual values" seems to have been ignored in the modification presented in your Item 10. If no actual drawings are to take place, what incentive does the assessor have to make careful assessments? And if this is so, why is this any improvement over the verbal approach? Of course, the assessor might take the matter seriously for a time; but eventually he would find that no drawings were to be made, and he might then lose interest. In this case (the

modification), it seems that the <u>threat</u> of actual drawings is used to provide incentive. Otherwise, this case could be extended to the case where no actual drawings are possible.

These comments have been scattered and not <u>directly</u> related to the actual questions posed in the note. Nevertheless, if you think they are of any value, feel free to circulate them in the "Round Robin." Since I am most interested in the questions posed (and left unanswered by me) I look forward to reading the comments of other participants.

10 August 1966   Indiana University, Bloomington

NOTES

1. On the Colloquium, now 14 years in existence, see J. Marschak (1972).

2. I follow here, in essence, a lecture delivered at Columbia University in 1950 and published in 1954. See Bibliography.

3. This is meant to apply to "stochastic" theories of choice -- see also p. 99 of Block and Marschak (1960); but it applies also to non-stochastic models.

4. Only finite sets of events have been considered here and, for example, in J. Marschak (1968), p. 49; (1970), Section 6; (1973); and J. Marschak and R. Radner (1972), Chapter II, Sections 8-11.

5. The procedure has been actually applied by Stael von Holstein (1970). See also Savage (1971).

BIBLIOGRAPHY

Bayes, Thomas. 1763. "An essay towards solving a problem in the doctrine of chances." The Philosophical Transactions, 53,370-418. Reprinted in Biometrika, 1958, 45, 296-315.

Bernoulli, Daniel. 1738. Specimen theoriae novae de mensura sortis. English in Econometrica, 1954, 22, 213-236.

Block, H.D. and J. Marschak. 1960. "Random orderings and stochastic theories of responses," p. 99, Contributions to Probability and Statistics in Honor of H. Hotelling, 97-132. Stanford.

Borch, Karl. 1968. The Economics of Uncertainty. Princeton University Press.

Borel, Emil. 1924. "À propos d'un traité de probabilités." Revue Philosophique, 98,321-336. English in Kyburg and Smokler, 1964.

Carnap, Rudolf. 1962. "The aim of inductive logic." Logic, Methodology, and Philosophy of Science, P. Suppes and A. Tarski, eds. Stanford.

De Finetti, Bruno. 1937. "La prévision: Ses lois logiques, ses sources subjectives." Annales de l'Institut H. Poincaré, 7, 1-63. Revised English language edition in Kyburg and Smokler, 1964.

—————. 1971. "Probabilità di una teoria e probabilità dei fatti." Studi di Probabilità ... in onore di G. Pompilj, 86-106. Gubbio.

————— and L.J. Savage. 1962. "Sul modo di scegliere le probabilità iniziale." Biblioteca del Metron, C, 1, 81-151. With English summary.

DeGroot, Morris H. 1970. Optimal Statistical Decisions. McGraw-Hill.

Ferguson, Thomas S. 1967. Mathematical Statistics: A Decision Theoretic Approach. Academic Press.

Good, Irving John. 1965. The Estimation of Probability: An Essay on Modern Bayesian Methods. M.I.T. Press.

—————. 1969. "A subjective evaluation of Bode's law and an objective test for approximate numerical rationality." J. Amer. Stat. Assn., 64, 23-66.

Heller, J. 1960. "Some numerical experiments for an M × J flow shop and its decision-theoretical aspects." Operations Research, 8, 176-184.

Kyburg, Henry E. and Howard E. Smokler. 1964. Studies in Subjective Probability. Wiley.

Marschak, Jacob. 1954. "Probability in the social sciences." Mathematical Thinking in the Social Sciences, P.F. Lazarsfeld, ed. The Free Press. Reissued, Russell & Russell, 1969.

—————. 1968. "Decision-making: Economic aspects." International Encyclopedia of the Social Sciences, 4, 42-55.

_____. 1970. "The economic man's logic." Induction, Growth and Trade. Essays in Honor of Sir Roy Harrod, 38-54. Clarendon Press.

_____. 1972. "Interdisciplinary discussions on mathematics in behavioral sciences." Festschrift fuer Walter Georg Waffenschmidt. Verlag Anton Hain.

_____. 1973. "Intersubjektive wahrscheinlichkeit." Heidelberger Jahrbuecher, 17,14-25. Springer.

_____ and Roy Radner. 1972. Economic Theory of Teams. Yale University Press.

Pratt, John W., Howard Raiffa and Robert Schaifer. 1965. Introduction to Statistical Decision Theory. McGraw-Hill.

Ramsey, Frank P. 1926-1928. The Foundations of Mathematics and Other Logical Essays. Reissued in 1950. New York: Humanities.

Savage, Leonard J. 1954. The Foundations of Statistics. Wiley. Revised edition, Dover Publications, 1972.

_____. 1971. "Elicitation of personal probabilities and expectations." J. Amer. Stat. Assn., 66,783-801.

Shohat, J.A. and J.D. Tamarkin. 1950. The Problem of Moments. New York: American Mathematical Society.

Stael von Holstein, Carl-Axel S. 1970. Assessment and Evaluation of Subjective Probability Distributions. Stockholm: Economic Research Institute.

Tversky, Amos. 1972. "Elimination by aspects: A theory of choice." Psychological Review, 79,281-299.

Winkler, Robert L. 1972. An Introduction to Bayesian Inference and Decision. Holt, Rinehart and Winston.